# The Battle of Subway Stops

Peilin Xin

July 23, 2020

# Introduction

- **Background**: Choosing a location for a restaurant is just one of a few keys to profitability. There is a famous saying in real estate that also works for opening a restaurant: “You make your money when you buy.”
- **Problem**: We will use the power of data science to generate a few viable locations to open a restaurant. The decision will be based on the locations accessibility, target market, population base and competition.
- **Interests**: This project aims to help those who are interested in opening an American restaurant in Boston and confused about where to open it.

# Data acquisition

- **Data sources**: To learn the accessibility of a location, I need to have subway and bus stops information which I can download from the **MBTA** (Massachusetts Bay Transportation Authority)'s official website. Then I will use **Foursquare API** to extract information of the restaurants in my target market around the transportation stops to learn the density of the restaurants.

# Data Cleaning

The most recent MBTA subway stops data has given the stops id, name, coordinates and a series of information and here are the data columns that I think could be useful for further analysis:

| | stop_name | stop_lat | stop_lon | zone_id | location_type | wheelchair_boarding | municipality | on_street | at_street |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Washington St opp Ruggles St | 42.330957 | -71.082754 | ExpressBus-Downtown | 0 | 1 | Boston | Washington Street | Ruggles Street |
| 1 | Theo Glynn Way @ Newmarket Sq | 42.330555 | -71.068787 | LocalBus | 0 | 1 | Boston | Theodore Glynn Way | Newmarket Square |
| 2 | Tremont St opp Temple Pl | 42.355692 | -71.062911 | LocalBus | 0 | 1 | Boston | Tremont Street | Temple Place |
| 3 | Albany St opp Randall St | 42.331591 | -71.076237 | LocalBus | 0 | 0 | Boston | Albany Street | Randall Street |
| 4 | Albany St opp E Concord St | 42.335017 | -71.071280 | LocalBus | 0 | 1 | Boston | Albany Street | East Concord Street |

Filling the empties with calculated values are not acceptable in this project because even a slight discrepancy with fact would be misleading. So, if there are missing data in this location, then this location will not be considered in the research process.

# Methodology

I used the Foursquare API to search for the restaurants that were located within 350 meters of each stop and extract the information such as venue's id, name, categories, coordinates, and distance from the stop. Here are some useful information of the dataset.

```
print('Average number of restaurants in each neighbor:', np.array([len(r) for r in restaurants]).mean())
```

```
Average number of restaurants in each neighbor: 2.532934131736527
```

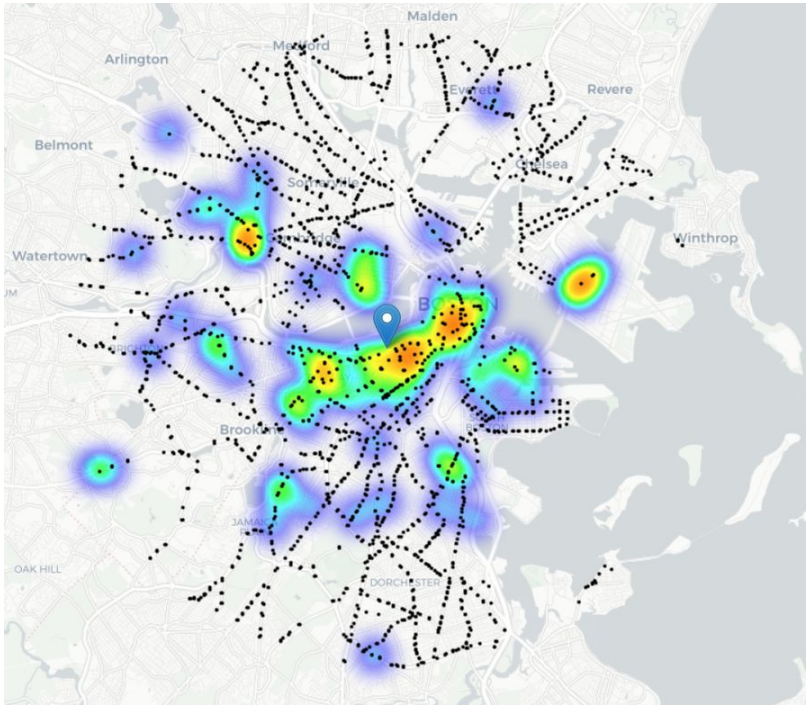
```
print('Maximum number of restaurants in a neighbor:', np.array([len(r) for r in restaurants]).max())
```

```
Maximum number of restaurants in a neighbor: 61
```

# Methodology

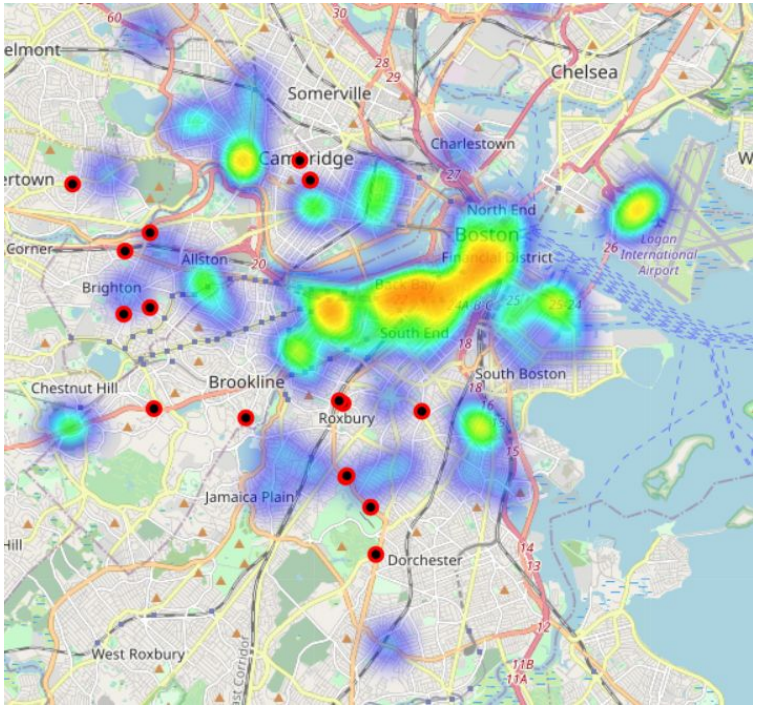

If we target the market to only American restaurants, we have 553 restaurants in the research range. Here is how the restaurants are distributed on the map.

# Methodology




If we take a look at the comparison between the heatmap of the restaurants and the whole transportation system, we can see that there are still a lot of stops that have potential.

# Result




I found out the locations with no more than one restaurant nearby and the locations with no American restaurants within 340 meters. These conditions either satisfied the low restaurant density area or the accessibility of the locations. When I inner-joined 2 conditions, I got 17 locations that met both.

# Result

```
----------------------------------------------------------------
Approximate American Restaurant Addresses recommended
----------------------------------------------------------------

American Legion Hwy @ Blue Hill Ave, Boston, MA 02124, USA
Market St @ Lothrop St, Boston, MA 02135, USA
Chestnut Hill Ave @ Jackson Ave, Boston, MA 02135, USA
N Beacon St opp Vineland St, Boston, MA 02109, USA
Columbus Ave @ Walnut Ave, Boston, MA 02119, USA
Centre St @ Cedar St, Boston, MA 02119, USA
Columbus Ave @ Cedar St, Boston, MA 02119, USA
Washington St @ Monastery Rd, Boston, MA 02135, USA
Cambridge St @ Dana St, Cambridge, MA 02138, USA
381 Dudley St opp Hampden St, Boston, MA 02119, USA
High St @ Cypress St, Brookline, MA 02445, USA
Boylston St @ Timon Ave, Brookline, MA 02467, USA
204 Seaver St, Boston, MA 02121, USA
Columbus Ave @ Walnut Ave, Boston, MA 02119, USA
Mt Auburn St @ Adams Ave, Watertown, MA 02472, USA
Broadway @ Lee St, Cambridge, MA 02139, USA
Broadway @ Fayette St, Cambridge, MA 02139, USA
```

After getting the potential locations that can be used to open a restaurant, I use Google API to geocode the actual addresses of these locations.

# Conclusion

The purpose of this project is to identify viable locations for opening an American restaurant. After the effort I have done, we had a general view of how the problem can be solved. The location generated can help us learn the places that can be potentially viable for opening a restaurant. However, we need more information if we want to push the analysis further.